Deformable Shape Detection and Description via Model-Based Region Grouping

Lifeng Liu and Stan Sclaroff[†]

Computer Science Department

Boston University

Boston, MA 02215

Abstract

A method for deformable shape detection and recognition is described. Deformable shape templates are used to partition the image into a globally consistent interpretation, determined in part by the minimum description length principle. Statistical shape models enforce the prior probabilities on global, parametric deformations for each object class. Once trained, the system autonomously segments deformed shapes from the background, while not merging them with adjacent objects or shadows. The formulation can be used to group image regions based on any image homogeneity predicate; e.g., texture, color, or motion. The recovered shape models can be used directly in object recognition. Experiments with color imagery are reported.

1 Introduction

Segmentation using traditional low-level image processing techniques, such as region growing, edge detection, and mathematical morphology operations, requires a considerable amount of interactive guidance in order to get satisfactory results. Automating these model-free approaches is difficult because of noise, shape complexity, illumination, inter-reflection, shadows, and variability within and across individual objects.

One can exploit prior knowledge to sufficiently constrain the segmentation problem. When available, such information can be used to eliminate ambiguities and reduce computational complexity in finding optimal groupings of image regions. For instance, model-based segmentation can be used in concert with image preprocessing to guide and constrain region grouping [13, 28, 35].

The use of models in segmentation is not a panacea, however. Due to shape deformation and variation within object classes, a simple rigid model-based approach will break down in general. This led to the use of deformable shape models in image segmentation [7, 18, 20, 22, 31, 38].

Another strategy is to utilize image features that are somewhat invariant to illumination [6, 16], or to directly model the physics of illumination, color, shadows, and sur-

face inter-reflections [14, 23]. Such approaches have been shown to improve segmentation accuracy, and could be combined with model based methods.

The above mentioned techniques make mistakes in merging regions, even in constrained contexts, because local constraints are in general insufficient. For more reliable segmentation, global consistency must be enforced. This idea is embodied in the *principle of global coherence* [33]: the best partitioning is the one that globally and consistently explains the greatest portion of the sensed data. Ideally, this should be coupled with the *minimum discription length (MDL) principle:* the simplest region segmentation explaining the observations is the best [11, 21, 24, 39].

Finding the globally consistent, MDL image labeling is impractical in general due to the computational complexity of global optimization algorithms. This has led to the use of parallel algorithms [11, 24] or approximation algorithms [5, 8, 15, 21, 29, 32, 37, 39].

2 Overview of Approach

The above mentioned work leads to the development of our approach. Deformable shape templates are used to partition the image into a globally consistent interpretation, determined in part by the MDL principle. The formulation can be used to group image regions based on any image homogeneity predicate; *e.g.,* texture, color, or motion.

Each shape template is specified in terms of global warping functions applied to a closed polygon. In the implementation, the prior distribution on global deformations for each shape is assumed Gaussian, and estimated using region segmentations provided in a training set. In our experiments, approximately 40 training images are needed to train a model. Once trained, the system autonomously segments deformed shapes from the background, while not merging them with adjacent objects or shadows.

We will now give a brief overview of the segmentation process as it is applied to find four bananas in the example image of Fig. 1(a). First, the input color image is oversegmented via standard region-merging algorithms [2, 9], as shown in Fig. 1(b). Using this over-segmentation, candidate regions for interesting objects are determined based on their color features [6].

Next an edge map is computed for the input image, as shown in Fig. 1(c). The edge map is used to constrain consideration of possible grouping hypotheses later in region merging. Notable edges and their strengths can be detected
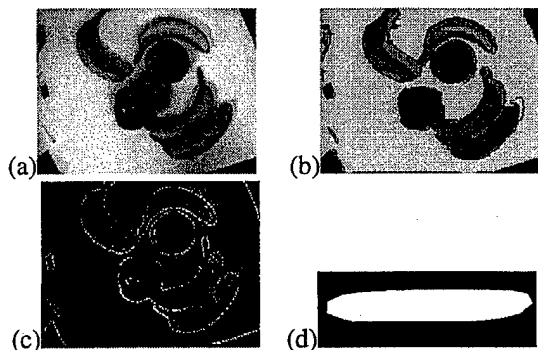
Figure 1: Example input and precomputation: (a) input image, (b) over segmentation, (c) edge map, (d) deformable template.
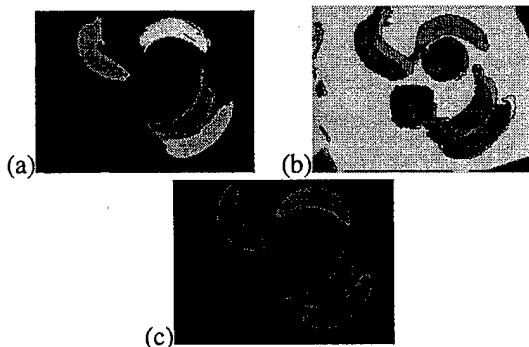

Figure 2: Result: (a) selected region groupings, (b) model-guided region merging, (c) recovered parametric shape models.

via standard image processing methods.

The system then tests various combinations of candidate region groupings. For each grouping hypothesis, we recover the model alignment and deformations needed to match the grouping. Fig. 1(d) shows the template used for grouping regions in this example. Goodness of fit is determined by a cost measure that includes: 1.) a region color compatibility term, 2.) a region/model area overlap term, and 3.) a deformation term. The third term enforces *a priori* constraints on the allowable deformations for a particular deformable shape class (*e.g.*, bananas). The template "prefers" to deform in ways that are consistent with the prior distribution on the deformation parameters.

In theory, the system should exhaustively test all possible combinations of regions groupings, and select the best ones for merging. In practice, region adjacency and edge map constraints are used to prune search. Despite this, the worst case computational complexity remains exponential. To make the problem tractable, we employ algorithms that find the approximately optimal solution: best-first, simulated annealing, or highest confidence first.

The approximately optimal region groupings obtained via the best-first algorithm are shown in Fig. 2(a). These groupings can then be merged in the color image segmentation, as shown in Fig. 2(b). Note that region merging

and object identification are executed simultaneously. The system simultaneously recovers a deformable template description for each region grouping as shown in Fig. 2(c). Recovered template parameters can be used in estimating the likelihood that a shape belongs to a particular class.

3 Related Work

Previous approaches are based on the active contours paradigm [22]. The snake formulation can be extended to include a term that enforces homogeneous properties over the region during region growing [7, 18, 20, 31, 38]. This hybrid approach offers the advantages of both region-based and deformable modeling techniques, and tends to be more robust with respect to model initialization and noisy data. However, it requires hand-placement of the initial model, or a user-specified seed point on the interior of the region. One proposed solution is to scatter many region seeds at random over the image, followed with segmentation guided via Bayes/MDL criteria [11, 39].

Other approaches use special-purpose deformable templates [19, 26, 38]; *e.g.*, to model facial features, such as eyes [38]. The template-based approach allows for inclusion of object-specific knowledge in the model. This further constrains segmentation, resulting in enhanced robustness to occlusion and noise. Under certain conditions, deformable templates can be derived semi-automatically, via statistical analysis of shape training data [10, 27]. The estimated probability density function (PDF) for the shape deformation parameters can be used in ML-estimation of segmentation and in Bayesian recognition methods.

From another view, image segmentation is a labeling problem; the ideal segmentation should be globally consistent or nearest to the one with maximum likelihood. This has led to various relaxation labeling or stochastic labeling methods that are related to general optimization algorithms [3, 17, 12]. Nearly all require some prior information, such as the number of labels needed or the probability distribution of labels in the image. Such information is not always available for general imagery.

After defining the criterion function for labeling, the next problem is computing the solution to the optimization problem. A number of proposed approaches employ simulated or deterministic annealing [5, 32, 15, 29, 37] (for a comparison see [25]). Chou and Brown [8] used highest confidence first (HCF) to infer a unique labeling from the posteriori distribution that is consistent with both the prior knowledge and evidence. Their method is analgous to deterministic annealing, but computation is more efficient.

A number of authors have proposed a formulation of the image partitioning problem that is based on the minimum description length (MDL) principle [11, 21, 24, 39]. MDL is based on information-theoretic arguments: the simplest model explaining the observations is the best. It also results in an objective function with no arbitrary thresholds.

2

As will be seen, the global cost function employed in our system is compatible with the MDL principle.

4 Deformable Model Formulation

In our system, a deformable model is used to guide grouping of image regions. Shape is specified in terms of global warping functions applied to a closed polygon, also known as a template. The global warping can be generic, and is determined by a vector of warping coefficients, $\mathbf{a}$. To demonstrate the approach, we implemented a system that uses quadratic polynomials to model global deformation due to scaling, shearing, bending, and tapering.

In a traditional active contours formulation, smoothness and bending operators are defined over the *control points* of the model to obtain a stiffness matrix, $\mathbf{K}$. In a deformable template formulation, we instead define a stiffness matrix over the *deformation parameters*. The strain energy is thus expressed in the template's deformation parameter space:

$$E_{strain} = \tilde{\mathbf{a}}^T \mathbf{K} \tilde{\mathbf{a}} \qquad (1)$$

where $\tilde{\mathbf{a}} = \mathbf{a} - \bar{\mathbf{a}}$ is a vector describing parameter displacement from a zero strain "rest" state.

There is a well understood link between active models and statistical estimation [10, 27, 36, 34]. Let us assume that the distribution on deformation parameters for a particular shape category can be modeled as a multi-dimensional normal distribution. The distribution is characterized by its mean $\bar{\mathbf{a}}$ and covariance matrix Σ. For a given deformation parameter vector $\mathbf{a}$, the sufficient statistic for characterizing likelihood is the Mahalanobis distance:

$$E_{deform} = \tilde{\mathbf{a}}^T \Sigma^{-1} \tilde{\mathbf{a}}, \qquad (2)$$

where $\tilde{\mathbf{a}} = \mathbf{a} - \bar{\mathbf{a}}$. Thus inverse covariance is essentially a "statistical stiffness matrix." As will be described, $\bar{\mathbf{a}}$, Σ are acquired via supervised learning.

An eigenvector transform is used to precondition problem by diagonalizing (decoupling) the stiffness matrix [30, 10]. This reduces the computational complexity of evaluating Eqs. 1 and 2 and improves the model's robustness to noise. During model fitting, deformations are recovered in the decoupled parameter space.

4.1 Model Fitting

One important step in the image partitioning procedure is to fit each region grouping hypothesis with deformable models from the object library. During segmentation, the shape model is deformed to match each grouping hypothesis $\mathbf{g}_i$ in such a way as to minimize a cost function:

$$E(\mathbf{g}_i) = E_{color} + \alpha E_{area} + \beta E_{deform}, \qquad (3)$$

where α and β are scalars that control the importance of the three terms. The color compatibility term E_{color} is simply the norm of color covariance matrix for pixels within the

current region grouping. The region/model area overlap term is computed via $E_{area} = \frac{S_G S_m}{S_c^2}$, where S_G is the area of the region grouping hypothesis, S_m is the area of the deformed model, and S_c is the common area between the regions and deformed model. By using the degree of overlap in our cost measure, we can avoid the problem of finding direct correspondence between landmark points, which is not easy in the presence of large deformations.

Various approaches to minimizing such a cost function have been suggested in the literature: graduated non-convexity [4], multi-grid approaches [36], and nonlinear programming methods [1]. In our system, we employ the downhill-simplex method because it requires only function evaluations, not derivatives. Though it is not very efficient in terms of the number of function evaluations that it requires, it is still suitable for our application since it is fully-automatic, and reliable. Due to space limitations, readers are referred to [25] for implementation details. The procedure is accelerated via a multiscale approach.

4.2 Model Training

In our current system, the template is defined by the operator as a polygonal model. During model training, a collection of training images are first over-segmented as described in the previous section. For each over-segmented image, a human operator is asked to mark candidate regions that belong to the same object. The system then uses downhill-simplex method to minimize the cost function in Eq. 3, thereby matching the template to the training regions in a particular image. This process is repeated for all images in the training set. As more training data is processed, the system can then semi-automate training. The system can take a "first guess" at the correct region grouping and present it to the operator for approval [25].

5 Automatic Image Segmentation

Once trained, the deformable model guides the grouping and merging of color regions. The process begins with over-segmentation of the input image. An edge map is also computed via standard image processing methods. Using this over-segmentation, candidate region groupings are determined based on the color band-rate feature [6].

Two major constraints are used in the selection of candidate groupings. The first constraint is a spatial constraint: every region in a grouping hypothesis should be adjacent to another region in the same group. The second constraint is a region boundary compatibility constraint: if the average edge strength along the boundary between two region exceeds a threshold, then the pair of images are marked as incompatible. Finally, the number of candidate groupings can be further reduced by considering only those that include at least one region with relatively large area.

Local constraints are insufficient for obtaining reliable segmentation. To gain more reliable segmentation, global

consistency must be enforced [33]. In the global consistency strategy, for any possible partitioning of the image, we compute a global cost for the whole configuration:

$$\mathcal{E} = \sum_{i=1}^{n} r_i E(\mathbf{g}_i) + \gamma n, \qquad (4)$$

where γ is a scalar, n is the number of the groupings in the current image partitioning, r_i is the ratio of ith group area to the total area, and $E(\mathbf{g}_i)$ is the cost function for the group $\mathbf{g}_i$ (Eq. 3). In our experiments, $\gamma = 0.04$.

The first term in Eq. 4 is the sum of the model compatibilty for every grouping in the image partition. The second term corresponds to the code length (number of models employed), and thereby enforces a minimum description length criterion, along the lines of [24].

5.1 Approximating the Optimal Solution

Eq. 4 does not exhibit the optimal substructure property required for solution via dynamic programming methods [25]. Furthermore, after the initial segmentation, the number of candidate regions is not small in general. We therefore implemented a number of approximation algorithms. Such algorithms tend to find a near-optimal partition within a reasonable number of steps.

One such algorithm, *best-first*, is greedy. It examines only the local cost of merging (Eq. 3) at each step. First, a list of all possible grouping hypotheses is generated as described above. Once all grouping hypotheses have been fitted with shape models, we then compare the merging cost of different grouping hypotheses, selecting the hypothesis with minimum model cost. If the cost is less than a threshold, then the regions are merged. Any hypotheses that include these merged regions are then eliminated from further consideration. If any unmerged grouping hypotheses remain, then we select the one with the minimum cost and repeat the procedure. If the cost exceeds the threshold or the hypothesis list is empty, then the procedure stops.

If the number of candidate regions in the over-segmented image is very large, the best-first strategy tends to be inefficient; it sometimes requires hours to segment an image on a standard workstation (SGI R5K Indy). This led us to explore approaches that approximately optimize global cost (Eq. 4). Due to space limitations, readers are directed to [25] for pseudocode and details of a simulated annealing solution. In our experiments, the convergence of the simulated annealing algorithm, while markedly better than best-first, is still slow. There is an inherent tradeoff between annealing schedule and correctness of result.

5.2 Highest Confidence First Algorithm

A deterministic algorithm, highest confidence first (HCF), can be used to improve convergence speed [8, 21]. The HCF algorithm as applied to our problem is as follows:

1. Initialize the region grouping configuration such that every region in the over-segmented image is in its own distinct group $\mathbf{g}_i$.

2. Fit models to each region grouping $\mathbf{g}_i$. Compute the global cost $\mathcal{E}_o$ via Eq. 4. Save this configuration as best found so far, C_o.

3. Set $\mathcal{E}_m$ to a very large value.

4. For each pair of adjacent groups $\mathbf{g}_i, \mathbf{g}_j$ in the current configuration, compute the global cost, $\mathcal{E}_2$ that would result if $\mathbf{g}_i, \mathbf{g}_j$ were merged. If $\mathcal{E}_2 < \mathcal{E}_m$, then set $\mathcal{E}_m = \mathcal{E}_2$ and save this merged configuration C_m. After this step, C_m is the configuration with minimum merging cost for any pair of groups in the current configuration.

5. Use the merged configuration C_m as the new configuration. If $\mathcal{E}_m < \mathcal{E}_o$, then set $\mathcal{E}_o = \mathcal{E}_m$ and save this new configuration as best found so far $C_o = C_m$.

6. Terminate when all groups are merged into one. and output the best configuration C_o and its cost value $\mathcal{E}_o$. Otherwise, go to 3.

In our experience, the computational complexity of HCF is generally less than that needed to obtain similar quality segmentation results via the simulated annealing algorithm [25]. In each HCF iteration, the number of different merging configurations tested is about $\mathbf{O}(n)$, where n is the number of regions in the over-segmented image. This is because some results from the previous iteration can be reused in the next. At each iteration (except the first), the algorithm need only compute the pairwise merging cost between all groups $\mathbf{g}_i$ and the newly-merged group from previous iteration. Thus the total complexity is $\mathbf{O}(n^2)$.

6 Examples

The system has been tested on hundreds of images from a number of different classes of cluttered color imagery: images of fruit, vegetables, and leaves collected under controlled lab conditions, and images of fish obtained from the world wide web. A few examples are now shown.

The first example shows results for detecting and merging regions associated with bananas. The shape template (Fig. 1(d)) was trained using 40 example images of bananas at varying orientations and scales. The training images were excluded from the test image data set. All images in the test data set were then segmented using the trained model, as described in Sec. 5. The best first strategy was employed in finding the best image partition.

The resulting model-based region groupings are shown below each of the original images in Fig. 3. In cases where there were multiple yellow objects in the image, the system recovered multiple model-based groupings (shown in different colors). Segmentation took between 30 and 180 sec. per image on an SGI R5K Indy workstation.

The system correctly grouped regions despite shadows, variation in illuminant, and shape deformation. Especially notable are cases where multiple yellow shapes abut each

Figure 4: Two deformable template models employed in our experiments: (a) leaf model, (b) fish model. The polygonal model was defined by the user, and then trained as described in Sec. 4.2.
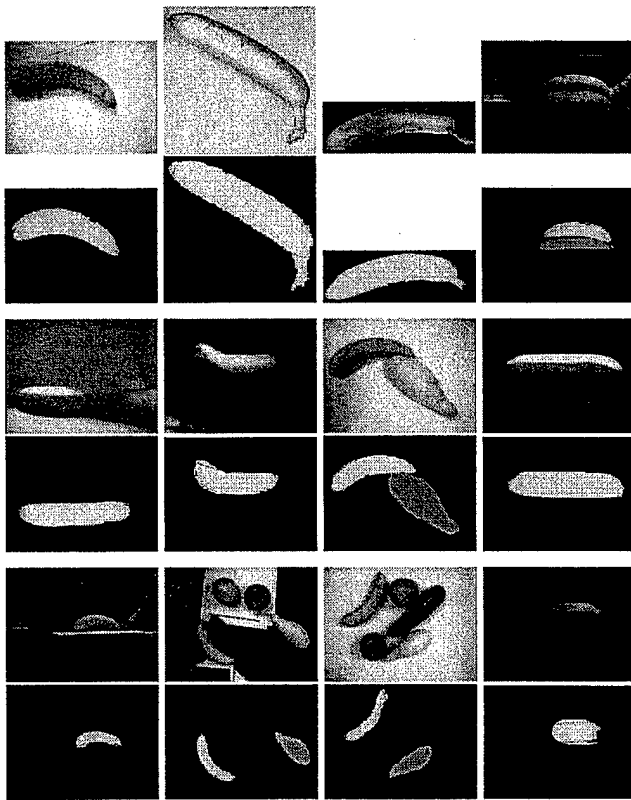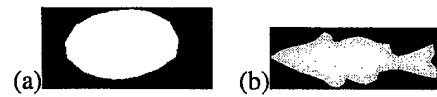

Figure 3: Image segmentation example: color images of bananas in various positions with varying illumination. The resulting model-based region groupings are shown below each color input image. If an image contained more than one detected shape, the shape that the system recognized as most "banana like" in each image is labeled in light gray. Note that the most similar shapes are other bent bananas of similar aspect ratio.

other. Due to the use of model-based region merging, the system is able to avoid merging similarly colored, adjacent but separate objects. The approach is also adept at avoiding merging objects with their similarly-colored shadows.

As explained in Sec. 4, each region grouping has an associated vector of shape deformation parameters a. The vector provides a low-dimensional description of each shape that can be stored and used for recognition. In cases where multiple objects are present, the system stores a list of model descriptions for that image.

Preliminary experiments in using the recovered shape parameter vectors for object recognition have been conducted. An example is shown in Fig. 3. The "target" shape was the banana in the first image (upper left). The subsequent images are shown in similarity ranking, left to right, top to bottom. Similarity was determined via Mahalonobis distance between recovered a vectors. The most similar shape in each image is shown highlighted in lighter gray in the labelled image below. The most similar shapes are other bent bananas of similar aspect ratio.

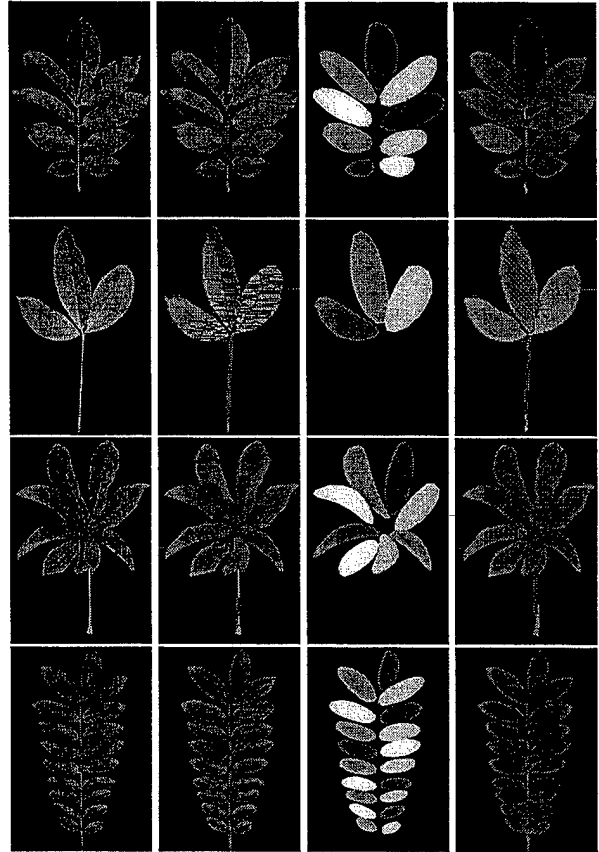The next example makes use of the global consistency

Figure 5: Leaf image segmentation examples. Each row of the figure shows one example. Original images are shown in the first column, followed by over-segmented images used as input to the merging algorithm. The third image in each row shows the best model configuration obtained via HCF. The model-based region merging result is shown as the final image in each row.

strategy to obtain segmentation of tropical leaf images. This example can be characterized by clutter of many simple leaves. The leaf model employed in this example was approximately an oval, as is shown in Fig. 4(a). It was defined and trained as in the previous example. The training images were not contained in our test image data set. The HCF algorithm was used in finding the "best" global configuration, as described in Sec. 5.2.

The method was tested on a collection of over 100 images of different tropical leaves. Due to space limitations, not all results can be shown here. Four examples are shown in Fig. 5. Segmentation took between 30 and 360 secs. per
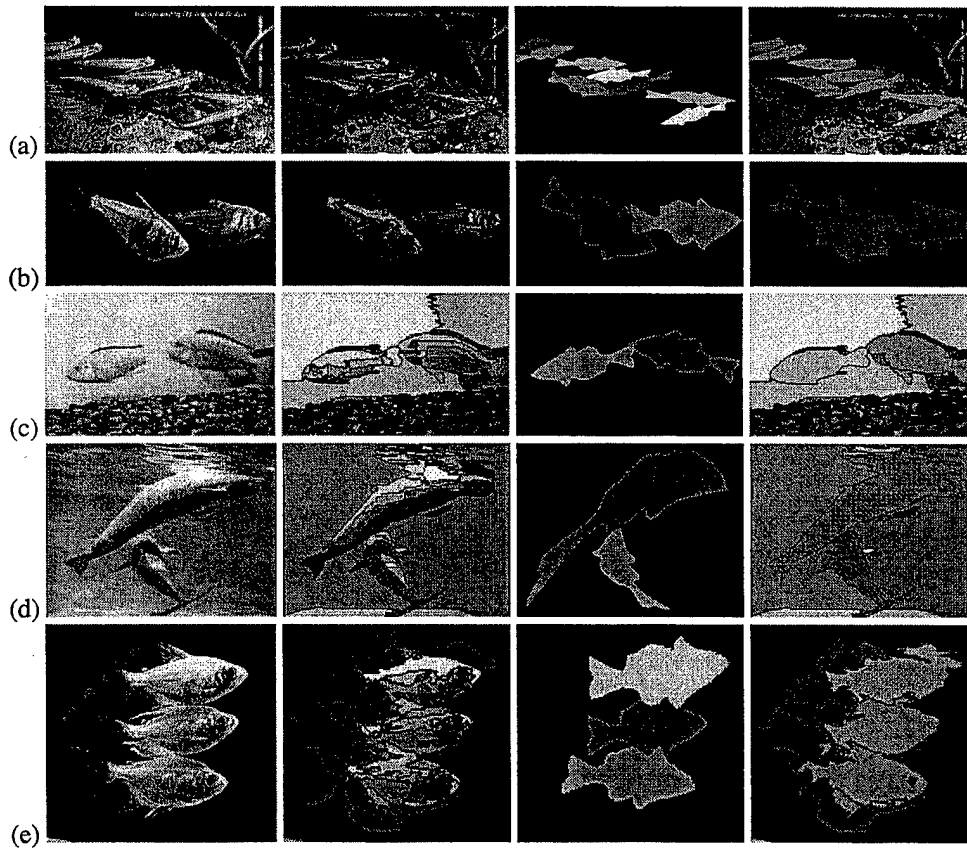
Figure 6: Example segmentation for images of fish. The original images are shown in the first column, followed by the over-segmented images used as input to the merging algorithm. The third column shows the models selected in the best merging configuration obtained via HCF. Finally, last column depicts the model-based merging.

image using the HCF algorithm. As can be seen, the system produces a satisfactory segmentation in each case, despite large deformations. Furthermore, the system does not merge adjacent, similarly colored regions *unless* they were consistent with the deformable shape model.

The final example shows segmentation results for five examples of fish images obtained from the world wide web. These images are particularly challenging, since there is greater shape and color variation, large deformation, and clutter. The fish model used in segmentation is shown in Fig. 4(b), and was trained using about 60 training images. The test images were excluded from the training set.

As shown in Fig. 6, the method recovered a deformable model description of each fish in the image. In one case, (Fig. 6(a)), the orientation of the model was incorrectly estimated for three fish. In such a case, local features might be used to resolve the orientation ambiguity. Despite clutter, large deformation, shape variation, and partial occlusions, the other fish were accurately segmented.

7 Discussion

In previous approaches to deformable template-based segmentation, initial model placement is either given by the operator, or obtained via exhaustively testing the model in all orientations, scales, and deformations centered at every pixel (or at random seed pixels). The region-based approach proposed in this paper significantly reduces the need to test all model positions.

Issues of computational complexity were addressed through the use of various constraints as was described in Sec. 5, and the use of multi-scale fitting. However, the complexity is still daunting in cluttered imagery and needs to be improved. The major issue is computation time required to obtain a segmentation result. This led to the evaluation of different methods for obtaining "optimal" region groupings. At present, the method is well-suited to applications where shape segmentation can be precomputed (*e.g.*, image databases indexing).

If there are shadows or partially overlapping objects in the image, then the best-first strategy can sometimes get a better result since it can select the most confident group to merge first, and avoid fitting spurious objects. Unfortunately, the computational complexity of best-first strategy prohibits application in general imagery.

Compared with the best first strategy, the simulated annealing approach offers a significant reduction in computational complexity. However, the degree of reduction in

complexity depends on the annealing schedule, and there is a trade-off between the robustness and the speed. Therefore, the global consistency strategy (via HCF) offers a reasonable compromise between speed and accuracy. It is therefore the preferred method.

The method is able to obtain a satisfactory segmentation despite clutter, variation in illuminant, shape deformation, *etc*. Based on the statistical shape model, the algorithm can detect the whole object correctly, while at the same time, avoid merging objects with background and shadow, or merging adjacent multiple objects. Region merging and object identification are executed simultaneously. Recovered shape parameters can be used directly in recognition.

References

[1] A. Amini, T. Weymouth, and R. Jain. Using dynamic programming for solving variational problems in vision. *PAMI*, 12(9):855–867, 1990.

[2] J.R. Beveridge, J.S. Griffith, R.R. Kohler, A.R. Hanson, and E.M. Riseman. Segmenting images using localized histograms and region merging. *IJCV*, 2(3):311–352, 1989.

[3] B. Bhanu and O. Faugeras. Shape matching of two-dimensional objects. *PAMI*, 6(2):137–156, 1984.

[4] A. Blake and A. Zisserman. *Visual Reconstruction*. M.I.T. Press, 1987.

[5] G. Bongiovanni and P. Crescenzi. Parallel simulated annealing for shape detection. *CVIU*, 61(1):60–69, 1995.

[6] M. Brill. Can color-space transformation improve color constancy other than von kries? *SPIE*, 1913:485–492, 1993.

[7] A. Chakraborty, L. Staib, and J. Duncan. Deformable boundary finding influenced by region homogeneity. *CVPR*, 1994.

[8] P. Chou and C. Brown. The theory and practice of Bayesian image labeling. *IJCV*, 4(3):185–210, 1990.

[9] D. Comaniciu and P. Meer. Robust analysis of feature spaces: Color image segmentation. *CVPR*, 1997.

[10] T.F. Cootes, A. Hill, C.J. Taylor, and J. Haslam. Use of active shape models for locating structure in medical images. *I&VC*, 12(6):355–365, 1994.

[11] T. Darrell and A. Pentland. Cooperative robust estimation using layers of support. *PAMI*, 17(5):474–487, 1995.

[12] O. Faugeras and M. Berthod. Improving consistency and reducing ambiguity in stochastic labeling: An optimization approach. *PAMI*, 3(4):412–424, 1981.

[13] P. Fua and A.J. Hanson. Using generic geometric models for intelligent shape extraction. *AAAI-87*, pp. 706–711, 1987.

[14] R. Gershon, A. Jepson, and J. Tsotsos. Ambient illumination and the determination of material changes. *JOSA-A*, 3(10):1700–1707, 1986.

[15] R. Grzeszczuk and D. Levin. Brownian strings: segmentating images with stochastically deformable contours. *PAMI*, 19(10):1100–1114, 1997.

[16] G. Healey. Segmenting images using normalized color. *IEEE Trans. on Sys., Man, and Cyber.*, 22(1):64–73, 1992.

[17] R. Hummel and S. Zucker. On the foundations of relaxation labeling processes. *PAMI*, 5(3):267–287, 1983.

[18] J. Ivins and J. Porrill. Active-region models for segmenting textures and colors. *I&VC*, 13(5):431–438, 1995.

[19] A.K. Jain, Y. Zhong, and S. Lakshmanan. Object matching using deformable templates. *PAMI*, 18(3):267–278, 1996.

[20] T. Jones and D. Metaxas. Segmentation using deformable models with affinity-based localization. *CVRMed*, 1997.

[21] T. Kanungo, B. Dom, W. Niblack, and D. Steele. A fast algorithm for MDL-based multi-band image segmentation. *CVPR*, 1994.

[22] M. Kass, A.P. Witkin, and D. Terzopoulos. Snakes: Active contour models. *IJCV*, 1(4):321–331, 1988.

[23] G. Klinker, S. Shafer, and T. Kanade. A physical approach to color image understanding. *IJCV*, 4(1):7–38, 1990.

[24] Y. Leclerc. Constructing simple and stable descriptions for image partitioning. *IJCV*, 3(1):73–102, 1989.

[25] L. Liu and S. Sclaroff. Deformable shape detection and description via model-based region grouping. BU CS TR-98-017, 1998.

[26] K.V. Mardia, W. Qian, D. Shah, and K.M.A. Desouza. Deformable template recognition of multiple occluded objects. *PAMI*, 19(9):1035–1042, 1997.

[27] J. Martin, A. Pentland, S. Sclaroff, and R. Kikinis. Characterization of neuropathological shape deformations. *PAMI*, 30(2):97-112, 1998.

[28] M. Nagao, T. Matsuyama, and Y. Ikeda. Region extraction and shape analysis in aerial photographs. *CGIP*, 10:195–223, 1979.

[29] D. Noll and W. Von Seelen. Object recognition by deterministic annealing. *I&VC*, 15(11):855–860, 1997.

[30] A. Pentland. Automatic extraction of deformable part models. *IJCV*, 4(2):107–126, 1990.

[31] R. Ronfard. Region-based strategies for active contour models. *IJCV*, 13(2):229–251, 1994.

[32] G. Storvik. Bayesian approach to dynamic contours through stochastic sampling and simulated annealing. *PAMI*, 16(10):976–986, 1994.

[33] T. Strat. *Natural Object Recognition*. 1992.

[34] R. Szeliski. *Bayesian Modeling of Uncertainty in Low-Level Vision*. Kluwer Academic Publishers, 1989.

[35] J. Tenenbaum and H. Barrow. Experiments in interpretation guided segmentation. *AI*, 8:241–274, 1977.

[36] D. Terzopoulos. Regularization of inverse visual problems involving discontinuities. *PAMI*, 8(4):413–424, 1986.

[37] J. Wang. Stochastic relaxation on partitions with connected components and its application to image segmentation. *PAMI*, 20(6):619–636, 1998.

[38] A. Yuille, D. Cohen, and P. Hallinan. Feature extraction from faces using deformable templates. *IJCV*, 8(2):99–111, 1992.

[39] S. Zhu and A. Yuille. Region competition: Unifying snakes, region growing, and bayes/MDL for multiband image segmentation. *PAMI*, 18(9):884–900, 1996.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE August 1999	3. REPORT TYPE AND DATES COVERED

4. TITLE AND SUBTITLE

Deformable Shape Detection and Description via Model-Based Region Grouping

5. FUNDING NUMBERS

G N00014-96-1-0661

6. AUTHOR(S)

Lifeng Liu and Stan Sclaroff

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Computer Science Department
Boston University
111 Cummington Street
Boston, MA 02215

8. PERFORMING ORGANIZATION REPORT NUMBER

sclaroff-ONR-TR98-017

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Department of the Navy
Office of Naval Research
Ballston Centre Tower One
800 North Quincy Street
Arlington, VA 22217-5660

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

A method for deformable shape detection and recognition is described. Deformable shape templates are used to partition the image into a globally consistent interpretation, determined in part by the minimum description length principle. Statistical shape models enforce the prior probabilities on global, parametric deformations for each object class. Once trained, the system autonomously segments deformed shapes from the background, while not merging them with adjacent objects or shadows. The formulation can be used to group image regions based on any image homogeneity predicate; e.g., texture, color, or motion. The recovered shape models can be used directly in object recognition. Experiments with color imagery are reported.

14. SUBJECT TERMS

Non-rigid shape description and recognition; image and video database modeling, content-based retrieval

15. NUMBER OF PAGES

7

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
unclassified	unclassified	unclassified	UL